Partial listing

US005394509A

United States Patent [19]

Winston

[11] Patent Number: 5,394,509

[45] Date of Patent: Feb. 28, 1995

[54] **DATA PROCESSING SYSTEM AND METHOD FOR SEARCHING FOR IMPROVED RESULTS FROM A PROCESS**

[76] Inventor: **Patrick H. Winston**, 88 Monument St., Concord, Mass. 01742

[21] Appl. No.: **861,068**

[22] Filed: **Mar. 31, 1992**

[51] **Int. Cl.**[6] **G06F 15/18**

[52] **U.S. Cl.** **395/13**

[58] **Field of Search** 395/13

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,697,242	9/1987	Holland et al.	395/13
4,935,877	6/1990	Koaz	395/13
4,961,152	10/1990	Davis	395/13
5,048,095	9/1991	Bhanu et al.	395/13
5,136,686	8/1992	Koza	395/13
5,140,530	8/1992	Guha et al.	395/13
5,148,513	9/1992	Koza et al.	395/13
5,222,192	6/1993	Shaefer	395/13

OTHER PUBLICATIONS

Kohonen, Teuvo. Content Addressable Memories. (Berlin: Springer-Verlag, 1987), pp. 19–27.

E. I. Chang et al., "Using Genetic Algorithms to Select and Create Features for Pattern Classification", Int'l Joint Conf. on Neural Networks, 17–21, Jun. 1990.

D. E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning", Addison-Wesley, 1989, pp. 10–14, 94–102.

M. Gordon, "Probabilistic and Genetic Algorithms For Document Retrieval", Communications of the ACM, vol. 31, No. 10, Oct. 1988, pp. 1208–1218.

Use of a Genetic Algoithm for Economic Optimization of a Manufacturing System; Tenga et al; Simulators V 1988 (The Society for Computer Simulation; 18–21 Apr. 1988; pp. 282–287.

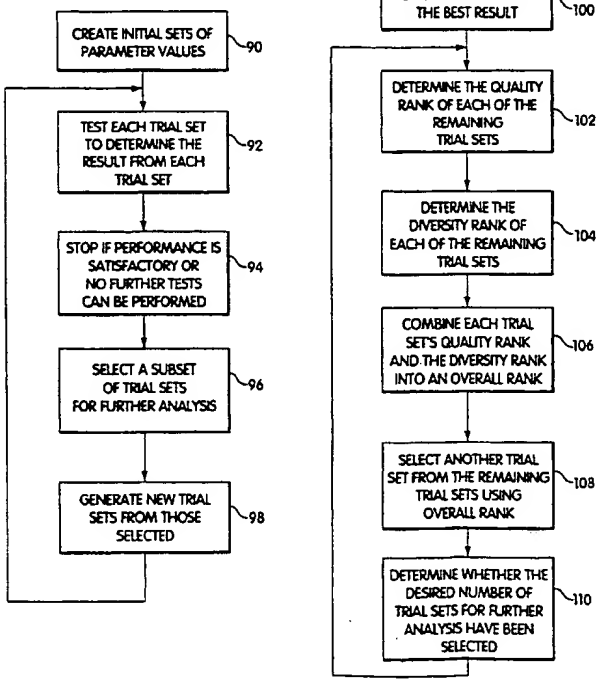
An Incremental Genetic Algorithm for Real-Time Optimization; Fogarty; Proc. of IEEE Int. Conf. on Systems,. Man, and Cybernetics; vol. 1; 14–17 Nov. 1989; pp. 321–326.

Primary Examiner—Allen R. MacDonald
Assistant Examiner—Robert W. Downs
Attorney, Agent, or Firm—Wolf, Greenfield & Sacks

[57] **ABSTRACT**

A data processing system and method for searching for improved results from the process utilizes genetic learning and optimization processes. The process is controlled according to a trial set of parameters. Trial sets are selected on the basis of an overall ranking based on results of the process as performed with a trial set. The ranking may be based on quality, or on a combination of rankings based on both quality and diversity. The data processing system and method are applicable to manufacturing processes, database search processes, and the design of products.

54 Claims, 7 Drawing Sheets



US-PAT-NO: 5394509

DOCUMENT-IDENTIFIER: US 5394509 A

TITLE: Data processing system and method for searching for improved results from a process

---------- KWIC ---------

Abstract Text - ABTX (1):

A data processing system and method for searching for improved results from the process utilizes genetic learning and optimization processes. The process is controlled according to a trial set of parameters. Trial sets are selected on the basis of an overall ranking based on results of the process as performed with a trial set. The ranking may be based on quality, or on a combination of rankings based on both quality and diversity. The data processing system and method are applicable to manufacturing processes, database **search** processes, and the design of products.

Brief Summary Text - BSTX (2):

The present invention is related to data processing systems and methods which assist in selection of parameters which control a process for the purpose of improving results obtained from the process. For example, the invention is related to selection of process parameters in a manufacturing process to improve the quantity or a quality of a product made by the process. The present invention is also related to database **search** methods and database systems for improving a prediction that an item in a database satisfies a predetermined selection criterion. The present invention is also related to design optimization processes. The data processing system and method of the present invention utilizes genetic learning and optimization processes.

Brief Summary Text - BSTX (5):

A database **search** involves similar problems. In this type of process, optimization methods may be used to improve a prediction as to whether an item in a database may satisfy some selected criterion. An item may include a number of characteristics. A **search** is performed using a number of sets of test characteristics, which are varied until a sufficient number of items which

match the test set satisfy the selected criterion. Those which do not match the test set should not satisfy the selected criterion.

Brief Summary Text - BSTX (6):

The range of possible results of a process combined with the range of possible parameters is known as the **search** space of the process. A difficult problem related to optimization methods is overcoming local maxima in the **search** space. This problem is related to the selection and generation of trial sets of parameters for the process. For instance, most optimization methods are "hill-climbing" methods which use small variations in the process parameters of known sets of parameters to generate new trial sets for each time the process is performed. When a local maximum is reached, a less than optimal result is obtained with such small variations to the process parameters on subsequent attempts. Thus, a local maximum may appear to be the optimal result, when, in fact, other maxima may exist. In an attempt to overcome this problem, most optimization or "hill-climbing" methods avoid known or discovered local maxima. Some methods are not capable of overcoming local maxima. Others may overcome local maxima, but require extensive experimentation and trials and often take an unacceptable length of time.

Brief Summary Text - BSTX (17):

It is another object of the invention to apply genetic learning processes to database **search** problems.

Brief Summary Text - BSTX (24):

Diversity among trial sets may be measured using a variety of well-known distance metrics. Each distance metric has advantages and disadvantages according to the **search** space of the optimization problem.

Detailed Description Text - DETX (8):

The diversity measure given above is only one of many possible diversity measures which may be used. The selection of a diversity measure is typically based on the **search** space, if it is known, in order to improve the accuracy of the diversity measure. Most diversity measures require that the measured items relate to a range of numerical values. Some process parameters may appear to be non-numeric, but could be translated into non-numeric values. For examples, colors (red, green, etc.) could be converted to light wavelengths. Addresses

could be converted to map coordinates. Diversity between non-numeric sets could be measured according to a Hamming distance. Hamming and Euclidean distance, along with other well-known diversity measures are described in Content Addressable Memories, second edition, by Teuvo Kohonen (Berlin: Springer-Verlag, 1987), pp. 19-27, the contents of which are hereby incorporated by reference.

Detailed Description Text - DETX (12):

Using diversity as a measure of fitness provides a different perspective on what may be done with local maxima in a **search** space, when taken in combination
with genetic processes such as mutation and crossover. Selection of trial sets to be crossed over may be performed on the basis of quality and diversity. Thus, this process of selection would suggest that many high quality and greatly diverse trial sets are preferable. The result is that known local maxima in the **search** space should be populated rather than avoided, in contrast
to other hill climbing or optimization methods.

Detailed Description Text - DETX (14):

Genetic learning algorithms may be adapted to include the foregoing constraints on selection of trial sets. Genetic learning processes such as these may be applied to processes such as manufacturing processes, database **searches** and design, in a manner to be described below, by using an appropriate
data processing system, such as shown in FIG. 3. The data processing system 59
includes a central processing unit 60 which controls the operation of the data processing system, including manipulation of data, and control of data flow. The data processing system includes a primary memory 62, which is typically volatile, such as a random access memory, and is used for temporarily storing data or application programs to be run by the data processing system. A secondary memory 64 is also used to provide permanent storage of data and application programs. Application programs include steps which are performed by the central processing unit 60 to complete a given process. The central processing unit 60 includes a program known as the operating system which controls data flow and execution of application programs. The data processing system 59 also preferably includes input devices 66 and output devices 68 which provide an interface to human operators. Such input devices 66 include keyboards, a mouse, voice recognition systems, and the like. Output devices 68 include video displays, printers, speech generation units, and the like. The data processing system 59 also may include a communication interface 70,

which
may include a modem and other appropriate communication application programs.
Such a communication interface 70 is useful for accessing remote computer systems. By using such a communication interface, a small computer such as an
IBM-PC.RTM. or a compatible machine, or an Apple.RTM. Macintosh.RTM. may be
used as the data processing system 59 unless the number and/or size of trial sets is large. Thus, larger computers, such as workstations, mainframes or supercomputers may also be used. Many important problems may require a mainframe-size or supercomputer for database testing or simulation. In general, any programmable general purpose computer or special purpose hardware
may be used.

Detailed Description Text - DETX (22):

Given a quality rank as determined in step 102 or a combined overall rank as determined in step 106, another trial set is selected from the remaining trial sets using the determined rank (step 108). The trial set having the highest rank may be selected or this selection could be performed probablistically according to a rank fitness formula such as equation 1 described above. In the examples described in the tables above, trial set A (1,4) would be selected as it has the highest overall rank. After another trial set is selected in step 108, it is then determined in step 110, whether the desired number of trial sets for further analysis have been selected. The desired number of trial sets may be a fixed number, or may be based on the number of known local maxima. In
some cases, the **search** space is full of local optima or maxima but those local optima tend to increase monotonically toward a global maxima. With such a **search** space, the number of "survivors" selected by step 108 may be periodically reduced and then allowed to increase again. Such periodic reduction in such a **search** space tends to eliminate trial sets stuck on low local maxima so they may be used to seek out higher local maxima.

Detailed Description Text - DETX (28):

A data processing system implementing the above-described genetic learning process may be used to improve, or optimize, many different specific processes. It is especially useful with database **search** applications, such as predictions using financial databases, and with manufacturing and design evaluation processes. How these implementations may be realized will now be described.

Detailed Description Text - DETX (30):

When this process is a database **search,** a trial set of parameters is typically a database query. With this process, a database is queried with a trial set to obtain a set of items from the database which match the trial set, and a set which do not match the trial set. It is then determined whether the matching and non-matching trial sets satisfy or fail to satisfy a given selection criterion. For example, a database of personal financial information could be searched with a database query which is intended to predict those people who are likely to go bankrupt. The results of this **search** could then be compared to information which determines whether in fact such individuals have gone bankrupt. Similarly, a trial set could be used to query a database with the intent of predicting whether certain individuals would be likely to buy a certain product. The matching and unmatching trial sets could be subjected to a market test, the results of which determine the quality to be assigned to the trial set. Another database application involves stockmarket prediction, where information concerning a company and its stock **price history** is stored in a database. A database query, intended to predict that a company will experience large growth, could be used to **search** the database. A comparison of the matched sets and unmatched sets to actual stock market prices would determine
the quality of the trial set as a predictor.

Claims Text - CLTX (36):

22. The method of claim 1, wherein the process is a database **search,** wherein the parameters are characteristics of an item stored in the database on which a **search** may be performed and wherein the result of the process performed
according to a trial set of parameters is

Claims Text - CLTX (75):

46. The data processing system of claim 25, wherein the process is a database **search,** wherein the parameters are characteristics of an item stored in the database on which a **search** may be performed and wherein the result of the process performed according to a trial set of parameters is

Other Reference Publication - OREF (3):

D. E. Goldberg, "Genetic Algorithms in **Search,** Optimization, and Machine Learning", Addison-Wesley, 1989, pp. 10-14, 94-102.